I am assuming that you are familiar with confidence intervals and some form of hypothesis testing. However, these topics can be taught from more than one perspective, and there are some common misconceptions regarding them, so it is worthwhile to give a review that will also lay a firm foundation for further work in statistics. I also want to introduce some notation that I will use in the course. So please read these notes carefully! They may contain details, perspectives, cautions, notation, etc. that you have not encountered before. I will, however, leave out some details that I assume you are familiar with -- such as the formulas for sample mean and sample standard deviation.

In statistics, we are studying data that we obtain as a sample from some population. (For example, we might be studying the population of all UT students and take a sample of 100 of those students.) We will assume that our sample is a simple random sample. That means that the sample is chosen by a method that gives every sample of the same size an equal chance of being chosen. (For example, we might choose our 100 UT students by assigning numbers to each UT student, and use a random number generator to pick a random sample of 100 numbers; our sample would then consist of the 100 students with those numbers.)

Typically we are interested in a random variable defined on the population under consideration. (For example, we might be interested in the height of UT students.) Typically we are interested in some parameter associated with this random variable. (For example, we might be interested in the mean height of UT students.) We will illustrate with the example of mean as our parameter of interest.

**Notation:** Let $y$ refer to the random variable. (e.g., height). Then
- The population mean (also called the expected value or expectation) of $y$ is denoted by either $E(y)$ or $\mu$.
- We use our sample to form the estimate $\bar{y}$ of $E(y)$.

More generally:
- We use the word parameter to refer to constants that have to do with the population. We will often refer to parameters using Greek letters (e.g., $\sigma$)
- We use the word statistic (singular) to refer to something calculated from the sample. So $\bar{y}$ is a statistic. (However, not all statistics are estimates of parameters.)
- I am using lower case letters to refer to random variables, since this is the notation used in the textbook. You might be (as I am) used to using capital letters to denote random variables and lower case to denote values of the random variable in a sample.
I will probably revert to that notation sometimes -- sometimes unconsciously, and sometimes to make the distinction clearer. I hope that the meaning will be clear from context.

**Models:** In order to use statistical inference or form confidence intervals, we need to have a *model* for our random variable. In the present context, this means we assume that the random variable has a certain (type of) distribution. Just what model (distribution) we choose depends on what we know about the random variable in question, including both theoretical considerations and available data. The choice of model is also usually influenced by information known about distributions -- we can deduce more from a distribution that has a lot known about it. In working with models (which we will do often in this course), always bear in mind the following quote from the statistician G.B.E. Box:

*All models are wrong - but some models are useful.*

For our example of height, we will use a normal model -- that is, we proceed under the assumption that the height of UT students is normally distributed, with mean $\mu$ and standard deviation $\sigma$. The values of $\mu$ or $\sigma$ are unknown; in fact, our aim is to try to use the data to say something about $\mu$. (If we are just considering students of one sex, both theory and empirical considerations indicate that a normal model should be a pretty good one; if we are considering both sexes, then data, theory, and common sense tell us that it isn't likely to be as good a choice as if we are just considering one sex. However, other theoretical considerations suggest that it probably isn't too bad.)

**Sampling Distributions:** Although we only have one sample in hand when we do statistics, *our reasoning will depend on thinking about all possible simple random samples of the same size* $n$. Each such sample has a sample mean $\bar{y}$, which is itself a random variable. (Note that the new random variable $\bar{y}$ depends on the choice of *sample*, whereas the original random variable $y$ depended on the choice of *student*.) Mathematics (using our assumption that the distribution of $y$ is normal with mean $\mu$ and standard deviation $\sigma$) tells us that the distribution of $\bar{y}$ is also normal with mean $\mu$, but its standard deviation is $\frac{\sigma}{\sqrt{n}}$. (Consequently, $\bar{y}$ varies less than $y$. See the demo Distribution of Mean at [http://www.kuleuven.ac.be/ucs/java/index.htm](http://www.kuleuven.ac.be/ucs/java/index.htm) under Basics) for an illustration of this. ) The distribution of $\bar{y}$ is called a *sampling distribution*. If we knew $\sigma$, we could use it to get a kind of margin of error for $\bar{y}$ as an estimate of $\mu$. Since we don't know $\sigma$, it is natural to use the *sample standard deviation* $s$ to estimate $\sigma$. (Note the
use of English letters to refer to the statistics, to distinguish them from the parameters, denoted by Greek letters.) However, using \( s \) instead of \( \sigma \) no longer yields a normal distribution. We can get around this difficulty by instead using the t-statistic 

\[
t = \frac{\bar{y} - \mu}{\text{se}(\bar{y})},
\]

where \( \text{se}(\bar{y}) = \frac{s}{\sqrt{n}} \) (the standard error of \( \bar{y} \)). This gives us still another random variable. Mathematical theory (plus our assumption of normality of \( y \)) tells us that this random variable \( t \) has a t-distribution with \( n-1 \) degrees of freedom.

Confidence Intervals: If we are trying to estimate \( E(y) \), we use a confidence interval to give us some sense of how good our estimate \( \bar{y} \) might be. (Note the qualifications in this sentence. Qualifications are important in statistics!) For a 95% confidence interval, we reason as follows: From tables or software, we can find the value \( t_0 \) of the t-statistic such that 2.5% of the area under the t-distribution (with \( n-1 \) degrees of freedom) lies to the right of \( t_0 \). Then in the language of probability,

\[
\Pr(-t_0 \leq \frac{\bar{y} - \mu}{\text{se}(\bar{y})} \leq t_0) = 0.95.
\]

Caution: In understanding this, it is important to remember that \( \bar{y} \) is our random variable, not \( \mu \). So this mathematical sentence should be interpreted as saying,

"The probability that a simple random sample of size \( n \) from the assumed distribution will produce a value of \( \bar{y} \) with \( -t_0 \leq \frac{\bar{y} - \mu}{\text{se}(\bar{y})} \leq t_0 \) is 95%"

With a little algebraic manipulation, we can see that this says the same thing as

\[
\Pr(\bar{y} - t_0\text{se}(\bar{y}) \leq \mu \leq \bar{y} + t_0\text{se}(\bar{y})) = 0.95.
\]

Bearing in mind the caution just mentioned, we can express this in words as,

"The probability that a simple random sample of size \( n \) from the assumed distribution will produce a value of \( \bar{y} \) with \( \bar{y} - t_0\text{se}(\bar{y}) \leq \mu \leq \bar{y} + t_0\text{se}(\bar{y}) \) is 95%.

The resulting interval \( (\bar{y} - t_0\text{se}(\bar{y}), \bar{y} + t_0\text{se}(\bar{y})) \) formed using the value of \( \bar{y} \) obtained from the data on hand is called a 95% confidence interval for \( \mu \). The confidence interval can be described in words in either of the following two ways:
i) The interval has been produced by a procedure that for 95% of all simple random samples of size \(n\) from the assumed distribution results in an interval containing \(\mu\).

ii) Either the confidence interval calculated from our sample contains \(\mu\), or our sample is one of the 5% of "bad" simple random samples of size \(n\) for which the resulting confidence interval doesn't contain \(\mu\).

(Of course, we also have to bear in mind the possibility that our assumed model is not a good one, or that our sample really is not a simple random sample.)

**Hypothesis tests:** We use a hypothesis test when we have some conjecture ("hypothesis") about the value of the parameter that we think might or might not be true. A hypothesis test is framed in terms of a null hypothesis, usually called \(H_0\) (or NH, as in the textbook). For all of the types of hypothesis tests we will do, the null hypothesis will be of the form

\[
\text{Parameter} = \text{specific value}.
\]

So in our example, where the parameter of interest is the mean, the null hypothesis would be stated as

\[
H_0 \text{ (or NH): } \mu = \mu_0.
\]

There are two frameworks for a hypothesis test. The one we will use in this course uses reasoning as follows: If the null hypothesis is true (and still assuming a normal model), then as above, we know that the sampling distribution of the statistic \(t = \frac{\bar{y} - \mu_0}{se(\bar{y})}\) (called the test statistic) has the t-distribution with \(n-1\) degrees of freedom. We calculate this test statistic for our sample of data (call the result of the calculation \(t_s\)), and then calculate the p-value, defined as the probability that a simple random sample of size \(n\) from our population would give a t-statistic at least as extreme as the one \((t_s)\) that we have calculated from the data, assuming the null hypothesis is true.

To pin down just what we mean by "at least as extreme," we usually specify an alternate hypothesis \(H_a\) (or AH, as in the textbook). This can be either two-sided or one-sided:

- **Two-sided:** \(H_a\) (or AH): \(\mu \neq \mu_0\)

- **One-sided:** This can take one of two forms -- either
If the alternate hypothesis is $\mu < \mu_0$, then "at least as extreme as" means $\leq$, so that the p-value is

$$p = \Pr(t \leq t_0).$$

Similarly, if the alternate hypothesis is $\mu > \mu_0$, then the p-value is

$$p = \Pr(t \geq t_0).$$

If the alternate hypothesis is two-sided, then the p-value is

$$P = \Pr(|t| \geq t_0).$$

The p-value is taken as a measure of the weight of evidence against $H_0$. A small $p$ means that it would be very unusual to obtain a test-statistic at least as extreme as ours if indeed the null hypothesis is true. Thus if we obtain a small $p$, then either we have an unusual sample, or the null hypothesis is false. (Or we don't have a simple random sample, or our model is not a good one.) We (somewhat subjectively, but based on what seems reasonable in the particular situation at hand) decide what value of $p$ is small enough for us to consider that our sample provides reasonable doubt against the null hypothesis; if $p$ is small enough to meet our criterion of reasonable doubt, then we say we reject the null hypothesis in favor of the alternate hypothesis.

**Note:**

1. A hypothesis test cannot prove a hypothesis. Therefore it is wrong to say, "the null hypothesis is false," or "the alternate hypothesis is true," or "the null hypothesis is true," or "the alternate hypothesis is false" on the basis of a hypothesis test.

2. Although it is arguably reasonable to say "we reject the null hypothesis" on the basis of a small p-value, there is not as sound an argument for saying "we accept the null hypothesis" on the basis of having a p-value that is not small enough to reject the null hypothesis. To see this, imagine a situation where you are doing two hypothesis tests, with null hypotheses just a little different from each other, using the same sample. It is very plausible that you can get a large (e.g., around 0.5) p-value for both hypothesis tests, so you haven't really got evidence to favor one null hypothesis over the other. So if your p-value is not small enough for rejection, all you can legitimately say is that the data are consistent with the null hypothesis. (This is assuming that by "accept" you mean that the data provides adequate evidence for the truth of the null hypothesis. If by "accept" you mean accept $\mu_0$ as a good enough approximation to the true $\mu$, than that's another matter -
- but if that's what you are interested in, using a confidence interval would probably be more straightforward than a hypothesis test.)

3. The p-value is, roughly speaking, the probability of getting data at least as extreme as the sample at hand, given that the null hypothesis is true. What many people really would like (and sometimes misinterpret the p-value as saying) is the probability that the null hypothesis is true, given the data we have. Bayesian analysis aims to get at the latter conditional probability, and for that reason is more appealing than classical statistics to many people. However, Bayesian analysis doesn't quite give what we'd like either, and is also often more difficult to carry out than classical statistical tests. Increasingly, people are using both kinds of analysis. I encourage you to take advantage of any opportunity you can to study some Bayesian analysis.

Many people set a criterion for determining what values of $p$ will be small enough to reject the null hypothesis. The upper bound for $p$ at which they will reject the null hypothesis is usually called $\alpha$. Thus if you set $\alpha = 0.05$ (a very common choice), then you are saying that you will reject the null hypothesis whenever $p < 0.05$. This means that if you took many, many simple random samples of size $n$ from this population, you would expect to falsely reject the null hypothesis 5% of the time -- that is, you'd be wrong about 5% of the time. For this reason, $\alpha$ is called the type I error rate.

Note:

1. If you set a type I error rate $\alpha$, then to be intellectually honest, you should do this before you calculate your p-value. Otherwise there is too much temptation to choose $\alpha$ based on what you would like to be true. In fact, it's a good idea to think about what p-values you are willing to accept as good evidence before the fact -- but if you are using p-values, you may think in terms of ranges of p-values that indicate "strong evidence," "moderate evidence," and "slight evidence," rather than just a reject/don't reject cut-off.

2. If you do set a type I error rate $\alpha$, then you don't really need to calculate $p$ to do your hypothesis test -- you can just reject whenever the calculated test statistic $t$ is more extreme ("more extreme" being determined as above by your alternate hypothesis) than $t_\alpha$, where $t_\alpha$ is the value of the t-distribution that would give p-value equal to $\alpha$.

3. If you are going to publish any scientific work, the second option is not a good choice; instead, you should calculate and publish the p-value, so others can decide if it satisfies their own criteria (which might be different from yours) for weight of evidence desired to reject the null hypothesis.

4. When an $\alpha$ has been chosen for determining when the null hypothesis will be rejected, and when the null hypothesis has indeed been rejected, many people say that
the result of the hypothesis test is "statistically significant at the \( \alpha \) level." It is important not to confuse "statistically significant" with "practically significant." For example, the improvement on a skill after a training session may be statistically significant, but could still be so small as to be irrelevant for practical purposes. By taking a large enough sample, almost anything can be shown to be statistically significant.

5. There is another variation of hypothesis testing which we will not use in this course. In this variation, you are trying to decide between two competing hypotheses, the null hypothesis \( H_0: \mu = \mu_0 \) and an alternate hypothesis \( H_a: \mu = \mu_a \) (still assuming we are testing means). Note that in this setting the alternate hypothesis specifies one value rather than being defined in terms of an inequality. Thus the null and alternate hypotheses play symmetric roles in the initial formulation of the problem. In this setting, you will either accept \( H_0 \) (and reject \( H_a \)) or accept \( H_a \) (and reject \( H_0 \)). You determine a rejection region. For values of the test statistic in the rejection region, you will reject the null hypothesis and accept the alternate hypothesis; otherwise you will accept the null hypothesis and reject the alternate hypothesis. In determining the rejection region, you take into account both the type I error rate \( \alpha \) and the type II error rate (the probability of accepting \( H_0 \) when \( H_a \) is true). Hypothesis tests of this sort are appropriate for situations such as industrial sampling. The costs of errors one way or the other as well as the costs of sampling are taken into account in determining the rejection region and the sample size.

Advice: Statistics gives many tools for obtaining information from data. However, it doesn't tell us "the answers." We need to combine what statistics tells us with careful thinking, caution, and common sense.